RAND

The Determinants of Private and Government Sector Earnings in Russia

Steven Stillman

DRU-2422

November 2000

Labor and Population Program
Working Paper Series 00–17

The Determinants of Private and Government Sector Earnings in Russia

Steven Stillman
RAND

Mail Stop M-15A
1700 Main Street
P.O. Box 2138
Santa Monica, CA 90407-2138
USA
Tel: 310-393-0411 ext. 7719
Fax: 310-451-7061
Email: stillman@rand.org

November 2000

I would like to thank Shelly Lundberg, Elaina Rose, and Judith Thornton for extremely helpful comments and suggestions. I am also grateful to seminar participants at the University of Washington for insightful discussions.

Abstract

So far little is known about how private sector earnings in Russia compare to those in the still strong government sector. This paper estimates sectoral earnings equations for rural and urban men and women which control for: (1) Self-selection into the workforce; and (2) Self-selection into either the private or government sector, while allowing for simultaneity in the selection decisions. The selection controls are found to have a considerable effect on the estimated sectoral earnings differentials for all four sample groups. Earnings differentials are examined by age, education, and unobserved skill. Expected earnings are found to be higher in the private sector for most groups.

Keywords: Sectoral Earnings, Private Sector, Russia, Heckman Selection Models

JEL-Code: J31, P23, C35

I) Introduction

In the past nine years, Russia has attempted to transform itself from a country whose economy is dominated by large state-owned firms to one whose economy is driven by private enterprise. While some effort has been made to promote individual and small business entrepreneurship, it has found it very difficult to develop a vibrant private sector.[1] Anecdotal evidence suggests that government sector firms and most larger enterprises, regardless of their ownership status, usually face soft-budget constraints and primarily undergo rent-seeking activities (Åslund 1997). As discussed in Lin & Tan (1999), these firms inherited major policy burdens from the Communist era and typically attempt to maximize employment instead of profits.[2] Meanwhile, newly formed private sector firms are thought to be simple profit maximizers. The disparity between the sectors in: (1) The objectives of firms; (2) The availability of non-pecuniary entitlements; and (3) Earnings variability and employment security, should cause earnings patterns and the returns to human capital to differ in each sector (Stillman 2000; Adamchik & Bedi 2000).

The increasing availability of individual-level data has encouraged researchers to examine the changing labor market in Russia during the transitional period.[3] However, so far no research known to the author has estimated wage or earnings equations for Russia which allow the returns

---

[1] Stillman (2000) calculates the private sector's share of total employment in Russia to be 27% in 1994-1996, rising to 30% in 1998. In comparison, Flanagan (1994) finds it to be 45% in Poland and 35% in Hungary in 1994, and the OECD (1998) reports it to be 69% in Sweden (the lowest in the OECD) and 87% in the US in 1997.

[2] Policy burdens include having to pay for employee housing, medical needs, and pensions, and having to employ redundant workers.

[3] For example, Brainerd (1999) analyzes changes in wage inequality between 1991 & 1994 and Ogloblin (1999) estimates changes in gender earnings differentials between 1994 & 1996.

to human capital to vary by sector and control for the endogeneity of sectoral choice.[4] Past papers have typically included in the earnings (wage) equation an indicator variable for whether an individual is employed in the private sector. This approach leads to two serious problems: (1) If sectoral choice is endogenously determined with sectoral earnings, all estimated returns to human capital are biased; and (2) If the relationship between human capital and earnings varies across sectors, inequality measures and earnings differentials which compare workers across sectors are incorrect.

This paper uses representative panel data from the Russian Longitudinal Monitoring Survey (RLMS) for the years 1994 – 1996 & 1998 to estimate unbiased earnings equations for both government and private sector workers. These earnings equations allow for the returns to human capital to vary across sectors and control for sectoral self-selection. Furthermore, since a sizeable percentage of both the male and female labor force are not actively employed, the econometric model also controls for individual self-selection into the workforce.[5] This paper focuses on two questions: (1) To what extent does controlling for the two forms of self-selection (and allowing different returns to human capital in each sector) affect the estimated sectoral earnings differentials for rural and urban men and women; and (2) How do sectoral earnings differentials vary by age, education, and unobserved skills?[6] Answering these questions can help

[4] This type of model, often known as a switching regression, has been used to examine sectoral wages or earnings in a variety of other countries (Germany: Dustmann & van Soest (1998), Côte d'Ivoire: van der Gaag & Vijverberg (1988), Peru: Stelcner et. al. (1989), Poland: Adamchik & Bedi (2000)).

[5] 30% (24%) of women (men) do not work in a given year, and 17% (12%) of women (men) never work in the sample period. Work is defined as having either worked positive hours or received some earnings in the past month.

[6] The sectoral earnings differential for a particular group is defined as the expected log earnings in the private sector

give us a better idea of which individuals benefit from increased private sector employment and which are vulnerable to decreases in the size of the government sector.

I find that properly controlling for self-selection has a considerable effect on the estimated sectoral earnings differentials for rural and urban men and women. Selection-corrected earnings differentials are found to be positive (i.e. expected earnings are higher in the private sector) for the average worker in each sample group besides rural women. University educated workers and workers in their twenties typically gain the most from private sector employment. For men, sectoral earnings differentials are found to be continuously decreasing in age in both rural and urban areas. On the other hand, older women in both areas typically have earnings differentials as large as women in their twenties. Overall, I find that the workers most vulnerable to the decreasing size of the government sector include rural women, rural men over thirty-five, and urban men over fifty. Among those most likely to gain from private sector development are rural men in their twenties, urban women, and younger urban men. For all groups, except urban men over fifty, individuals with university degrees benefit more from private sector employment than other comparable workers.

This paper continues as following. Section II describes the data used in the analysis and presents summary statistics by work status and employment sector. Section III outlines the econometric model and the estimation strategy. Section IV presents the estimation results and discusses the effect that the selection controls have on them. Section V examines the sectoral differentials for various groups of individuals. Section VI then offers conclusions.

---

minus those in the government sector for the mean individual in that group. Expected earnings are based on the systematic parts of the estimated earnings equations and are unconditional on actual work status.

II) Data & Summary Statistics

A) The Data

All empirical work in this paper uses data from phase two of the Russian Longitudinal Monitoring Survey (RLMS) for the years 1994 – 1996 & 1998 (Rounds V – VIII), which is a household-based representative survey of Russia collected by the Population Center at the University of North Carolina.[7] The survey is designed as a repeated sample of each household dwelling. Thus, instead of following individuals or households from one year to the next, RLMS merely returns to the same dwelling sampled in the previous year. Consequently, by definition all households who move locally or migrate to another region are automatically lost to follow-up. Adding an additional twist to the sampling process, if the previous occupants of a sample dwelling are lost to follow-up, the new occupants are invited to join the survey sample.[8] In each year, data is collected at the individual, household, and community level.[9] The dataset used in all analyses includes all prime-age individuals, defined as men aged 18 to 59 and women aged 18 to 54 (considered the normal working ages for Russian men and women). After children and elderly are dropped from the dataset, the final sample consists of 9,239 individuals from 4,266 households, providing a total of 21,120 observations.[10]

Individuals are asked to report both monetary and in-kind earnings collected, and hours

---

[7] The project description at www.cpc.unc.edu/rlms provides complete information about the RLMS survey and its sampling procedure.

[8] Heeringa (1997) provides additional information on attrition in RLMS and discusses its overall representiveness.

[9] All individuals in each household are surveyed with the exception of some elderly and very young members. Extensive data is collected for each of the 160 communities (sites). Information is provided to assign the 160 sites to 38 primary sampling units (PSUs) and to 8 regions.

worked during the last month at their primary, secondary, and other (self-employment) places of employment. An individual's main job is defined as the place of employment where they had the highest total earnings or worked the most hours if earnings were equal. All analyses in this paper use only the data for this main job. Only individuals who had zero earnings and worked zero hours in all three job types are considered non-working. Individuals are also asked detailed information about each job. This information is used to classify the sector of employment at their main job. This paper's focus is on comparing individual earnings in the profit-maximizing sectors in Russia (small private enterprise and self-employment) to those in the non-competitive sectors (large private enterprise and the government). Thus, it uses a relatively strict definition of private sector employment, excluding all firms with partial government ownership or more than 200 employees. Appendix B provides additional information on how each worker's sector of employment is identified.

Currently, only price level data for metropolitan Moscow & St. Petersburg is available from international sources. However, it is possible to use RLMS to calculate a regional price index. As part of the community surveys, a local citizen is sent out to collect data on the prices of a selection of store bought and market goods. Households are also asked to provide information on their expenditures in the week prior to the survey. This information is used to derive a regional price index which has 1998 Moscow & St. Petersburg as the base region-year.[11]

B) Summary Statistics by Work Status & Sector

Table 1 presents summary statistics by work status and sector of employment. Individuals who

[10] Summary statistics for all variables are available in appendix A.

[11] Detailed information on the construction of this price index is available from the author by request.

do not work are more likely to be female, younger, and unmarried, to have lower levels of schooling, and to live in rural communities. Large regional differences in participation are also found with individuals in Moscow / St. Petersburg having high levels of participation, while individuals in the North Caucasus have low levels. Participation also appears to vary by community characteristics such as the availability of a local social welfare office or an employment service. As anticipated, individuals in communities with these local services are more frequently employed. Overall, the summary statistics strongly suggest that the sample of observed workers is a not a random sample of the overall population and thus ordinary least squares estimates of human capital returns will be affected by sample selection bias.[12]

Workers in the two sectors are quite similar across a broad range of characteristics. This is perhaps surprising considering that the average private sector worker earns 38% more than the average government sector worker. On the whole, private sector workers are more likely to be male, unmarried, and slightly younger, and to have less formal education than government sector workers. Again, large regional variation is found with workers in Moscow / St. Petersburg and North Caucasian more typically employed in the private sector, and workers in Volga-Vaytski / Volga Basin in the government sector. Private sector workers are more likely to own a home, to own more land, and to have higher amounts of real household assets. Stillman (2000) finds earnings variability to be higher in the private sector in each region using the same sample of Russian workers and suggests that individuals with greater assets are more likely to be employed in the private sector because they have greater ability to intertemporally smooth consumption. Household assets may also facilitate private sector employment if they are sold/used to pay for

[12] The direction and magnitude of these biases cannot be analytically determined, and thus the empirical results must be relied upon to ascertain their importance.

some of the fixed costs faced by entrants to the private sector (especially to self-employment). On the other hand, since individuals who work in the private sector receive higher earnings, on average, they should be able to accumulate greater assets than government sector workers. Private sector workers are also more likely to reside in communities with stronger banking infrastructure. Banking infrastructure, like household assets, may contribute to an individual's ability to smooth consumption or may provide startup capital for private sector businesses. Communities with stronger banking infrastructure may also have relatively better business environments.

III) Econometric Model & Estimation Strategy

A) The Sector-Specific Earnings Equations

As in most traditional earnings models, log earnings is modeled as a linear function of observable determinants of human capital. As discussed, the returns to human capital are likely to differ across sectors. Thus, earnings equations are estimated separately for each sector. The earnings equations are defined as following:

$$\ln Y_{it}^{S} = \beta_0^{S} + X_{it}\beta_1^{S} + u_t^{S} + u_{it}^{S}, \tag{1}$$

where S = gv (government) or pr (private), $\ln Y_{it}^{S}$ is log real earnings at individual i's primary job in the month prior to interview in year t and sector S, and X_{it} is a vector of individual characteristics (including gender, height, weight, age, marital status, educational status) and residential location variables (including region, urbanicity, and city status), which may influence an individual's marginal productivity in either sector.[13] The error term is decomposed into two

[13] Educational status includes indicator variables for: (i) Some General Secondary Education (U.S. High School); (ii) Complete General Secondary Education (normal academic standard); (iii) Ordinary Vocational Diploma (lowest

parts: u_t^S, which is a fixed sector-year component to account for possible business cycle effects, and u_{it}^S, which is an idiosyncratic component and is assumed to be normally distributed mean zero and variance σ_u^S. The fixed sector-year component is estimated using year dummy variables which are allowed to be correlated with X_{it}.

B) The Reduced Form Selection Equations

This paper does not attempt to explicitly model the individual's participation decision and instead relies upon a reduced form model of behavior. An individual decides whether or not to work by comparing their expected sectoral earnings to their reservation earnings. Thus, all individual, household, and community characteristics which may influence either expected or reservation earnings are included as independent variables in the reduced form participation equation.[14] Defining W_{it}^* as the unobservable index function underlying individual i's decision whether to work at time t, the reduced form equation is specified as

---

academic standard); (iv) Secondary Vocational Diploma (low academic standard); (v) Specialized Secondary Diploma (highest academic standard); (vi) Professional Course Diploma; and (vii) Institute / University Diploma. The eight regions represented are Moscow / St. Petersburg, Northern / North Western, Central / Central Black-Earth, Volga-Vaytski / Volga Basin (default), North Caucasian, Ural, West Siberian, and East Siberian / Far Eastern. City status includes indicator variables for: (i) Capital (Moscow) / Oblast Center (regional capital); (ii) Raion Center (major town); and (iii) Other (default).

[14] The two reduced form selection equations are jointly estimated thus controlling for unobserved factors which influence both participation and sectoral choice, and observed factors which indirectly affect the participation decision through their effect on sectoral choice.

$$W_{it}^{*} = \gamma_0 + Z_{1it}\gamma_1 + e_{1it},$$
$$W_{it} = 1 \quad \text{if } W_{it}^{*} > 0, \tag{2}$$
$$W_{it} = 0 \quad \text{if } W_{it}^{*} \leq 0,$$

where $W_{it} = 1$ if individual i is employed at time t and equals zero otherwise, and Z_{1it} is a vector of individual characteristics (as above with the addition of the year), household demographics (including the number of children, elderly, other adults, and other workers), household characteristics (including whether a home is owned, the amount of land owned, the value of real household assets, and the amount of other household income), community employment variables (including whether there is an employment service, whether government enterprises have been closed, and whether there is social welfare office), and residential location variables.[15] The error term, e_{1it}, is assumed to be normally distributed mean zero and variance normalized to one.[16]

An individual decides in which sector to participate by comparing their expected utility in each sector. Following Stillman (2000), individual sectoral labor supply is a function of: (1) The

[15] The value of real household assets totals the estimated worth of the following assets: (1) Refrigerator; (2) Separate Freezer; (3) Washing Machine; (4) Black & White TV; (5) Color TV; (6) VCR; (7) Car or Truck; (8) Motorcycle or Motorboat; (9) Tractor; (10) Garden Cottage; (11) Dacha (vacation home) or Other House; and (12) Other Apartment. The amount of other household income totals elderly pension income and household farm income in the previous month. Detailed information on the construction of these variables is available by request from the author. An employment service may perform any of the following functions: (1) Help people find work; (2) Conduct retraining in a new specialization; and/or (3) Pay unemployment benefits. A welfare office may: (1) Pay pensions; (2) Help people buy food; (3) Help people with housework or food preparation; and/or (4) Help people obtain medicines or medical services.

[16] The discrete choice models in this paper will be analyzed using maximum likelihood probit and bivariate probit estimation which requires that the variance of the error term be normalized.

difference in individual expected sectoral earnings; (2) The difference in regional sectoral earnings variability, where the effect variability has on sectoral choice is a function of characteristics which may influence individual ex-post consumption smoothing ability; and (3) Characteristics which may be correlated with individual preferences towards a particular sector. This paper does not estimate the structural sectoral labor supply model from which these relationships are derived.[17] Instead, it estimates the reduced form sectoral choice equation which includes as explanatory variables all exogenous variables which may influence sectoral choice through any of the above channels. Defining P_{it}^{*} as the unobservable index function underlying individual i's decision whether to work in the private sector at time t, the reduced form sectoral choice equation is specified as

$$\begin{aligned} P_{it}^{*} &= \psi_0 + Z_{2it}\psi_1 + e_{2it}, \\ P_{it} &= 1 \quad \text{if } P_{it}^{*} > 0, \\ P_{it} &= 0 \quad \text{if } P_{it}^{*} \leq 0, \end{aligned} \tag{3}$$

where P_{it} = 1 (0) if individual i is employed in the private (government) sector at time t, and Z_{2it} is a vector of individual characteristics, household demographics, household characteristics, community characteristics (including whether individuals can own private land, the availability of formal banking, and other bank characteristics), and residential location variables.[18] The error

[17] Stillman (2000) provides a description of the theoretical model which underlies the empirical setup in this paper and estimates the structural sectoral labor supply model.

[18] The availability of formal banking includes indicator variables for: (i) No Banks; (ii) Only branch office of the Russian Federation Savings Bank (default); and (iii) Banks or branches besides RFSB. Other bank characteristics includes indicator variables for: (i): Depositor can withdraw 1,000 rubles out of their savings account by the next day; and (ii) Nearest non – RFSB bank or branch office is < 1 mile away.

term, e_{2it}, is assumed to be normally distributed mean zero and variance normalized to one.

C) Estimating the Reduced Form Model

Gathering equations (1) - (3), combining the model coefficients and the independent variables each into a single vector, and defining $y_{it}^{S} = \ln Y_{it}^{S}$, the full reduced form model can be written as

$$\begin{aligned} W_{it}^{*} &= \mathrm{Z}_{1it}^{*}\gamma + e_{1it}, \\ P_{it}^{*} &= \mathrm{Z}_{2it}^{*}\psi + e_{2it}, \\ y_{it}^{gv} &= X_{it}^{*}\beta^{gv} + u_{it}^{gv}, \\ y_{it}^{pr} &= X_{it}^{*}\beta^{pr} + u_{it}^{pr}, \end{aligned} \tag{4}$$

where y_{it}^{gv} is observed for individual i at time t if $W_{it}^{*} > 0$ and $P_{it}^{*} \leq 0$ ($W_{it} = 1$ & $P_{it} = 0$), y_{it}^{pr} is observed for individual i at time t if $W_{it}^{*} > 0$ and $P_{it}^{*} > 0$ ($W_{it} = 1$ & $P_{it} = 1$), and no measure of earnings in available if $W_{it}^{*} \leq 0$ ($W_{it} = 0$). Denote cov (e_{1it}, e_{2it}, u_{it}^{gv}, u_{it}^{pr}) = Σ, where

$$\Sigma = \begin{bmatrix} 1 & \rho_e & \rho_1^{gv}\sigma_u^{gv} & \rho_1^{pr}\sigma_u^{pr} \\ \rho_e & 1 & \rho_2^{gv}\sigma_u^{gv} & \rho_2^{pr}\sigma_u^{pr} \\ \rho_1^{gv}\sigma_u^{gv} & \rho_2^{gv}\sigma_u^{gv} & \sigma_u^{gv} & - \\ \rho_1^{pr}\sigma_u^{pr} & \rho_2^{pr}\sigma_u^{pr} & - & \sigma_u^{pr} \end{bmatrix}. \tag{5}$$

For identification proposes it is necessary to assume that e_{1it} & e_{2it} are distributed bivariate normal, but it is unnecessary to restrict the distribution of the other terms in Σ.[19] The covariance between (u_{it}^{gv}, u_{it}^{pr}) cannot be identified since individuals are never simultaneously observed in both sectors. This model is estimated using a two-step procedure which first jointly estimates the two reduced form selection equations (2) & (3) using maximum likelihood bivariate probit

[19] Tunali (1986) provides additional information on the structure and identification of the two-step model estimated in this paper.

estimation and then estimates the earnings equations (1) controlling for the selection in the first step. Correct standard errors for this estimation procedure are generated via bootstrapping. Complete details on the estimation procedure, including the bootstrap method, are available in appendix C.

D) Identification of the Earnings & Reduced Form Selection Equations

Table 2 summarizes the identification of reduced form model. The selection terms and sector-specific earnings equations are readily identified since both of the selection equations include groups of variables which are uniquely excluded from the earnings equations. Community employment variables are included in the reduced form participation equation as they are assumed to affect an individual's reservation earnings; Individuals in communities which have: (1) An employment service; (2) A social welfare office; and/or (3) Not had government enterprises closed, are expected to have higher reservation earnings as an employment service should lower employment search costs for individuals, a social welfare office should decrease the cost to individuals of being unemployed, and not having government enterprises closed should lower the number of competing unemployed searching for jobs. These variables are excluded from both the earnings and the reduced form sectoral choice equations as they should not affect marginal productivity, ex-post consumption smoothing ability, or individual preferences towards a particular sector.

Community characteristics are included in the reduced form sectoral choice equation as they are assumed to affect an individual's ex-post consumption smoothing ability; Individuals in communities which: (1) Allow individuals to own private land; (2) Have local banks; and/or (3) Have high quality banks, are expected to have greater ability to intertemporally smooth consumption as financial markets can be used to lend and borrow, and land markets to (de-)accumulate assets. These variables are excluded from both the earnings and the reduced form

participation equations as they should not affect marginal productivity, expected earnings, or reservation earnings.

IV) Estimation Results

A) The Importance of Properly Controlling for Endogenous Selection

The reduced form selection equations (2) & (3) and the selection-corrected sector-specific earnings equations (1) are estimated separately for rural and urban men and women, and include a quintic in age as well as the variables noted.[20] Data on individuals are pooled across time-periods and the panel nature of the data is not exploited in these estimates. Six different specifications of the reduced form model are considered. These specifications impose different restrictions on the covariance matrix of the error terms and on the returns to human capital across the two sectors.

Table 3 demonstrates how these various restrictions affect sectoral log earnings differentials estimated for average workers across three education groups and for overall average workers, and displays the estimated error covariance matrix of each reduced form model.[21] In

[20] Earning are imputed for individuals who report zero earnings and work positive hours at their main job. Using the available data on hours worked at main job, wage rates are calculated for all individuals who work positive hours and receive positive earnings. The wage rate at the 1st percentile in a particular sector (multiplied by actual hours worked) is then used to impute earnings for all individuals with zero earnings in that sector. This is done to avoid biasing the sectoral earnings equations by leaving out these zero earnings workers who thus have missing log earnings.

[21] The sectoral earnings differential for group i is defined in all cases as the expected log earnings in the private sector minus those in the government sector for the mean individual in group i. Expected earnings are based on the systematic parts of the earnings equations and are unconditional on the actual work status of individuals in a particular group.

specification 1, returns to human capital are constrained to be equal across the two sectors and the possible endogeneity of both participation and sectoral choice is ignored. Specification 2 allows returns to vary completely across sectors, while still ignoring both endogeneity problems. Specification 3 extends the previous specification by controlling for the endogeneity of participation while ignoring sectoral choice. Specification 4 is the reverse of 3, controlling for the endogeneity of sectoral choice, but ignoring participation. Specification 5 controls for both potential endogeneity problems, but assumes the selection decisions are uncorrelated (i.e. individuals first choose whether to work and then choose in which sector to work). Specification 6 jointly estimates the two reduced form decision rules. This specification is the most general and all others are nested within it.[22]

The estimated sectoral earnings differentials are strongly influenced by the addition of the selection controls. For all groups, except rural men, failing to jointly control for both types of self-selection leads to extremely misleading conclusions.[23] Significant negative self-selection of participants into the government sector is found for rural women. This causes the sample earnings of rural women in the government sector to be much lower than the earnings an average rural women would receive in this sector. Accordingly, the selection-corrected sectoral earnings differential (-16.5 log percentage points) is much smaller than the uncorrected earnings gap (25.3) for rural women. For urban women and men, negative (positive) self-selection of workers

[22] The complete results from specifications 1 - 5 of the sector-specific earnings equations and from the reduced form selection equations are available by request from the author.

[23] The results for rural women appear to be unbiased when controlling for only endogenous participation, but the strong correlation found between participation and sectoral choice for this group suggests that it is better to control jointly for both types of self-selection.

into the private (government) sector is found. This causes the sample earnings of private (government) sector workers to be lower (higher) than the earnings an average worker would receive in this sector. Thus, for both urban women and men, the selection-corrected sectoral earnings differential (97.3 for women; 97.2 for men) is much larger than the uncorrected earnings gap (30.6; 20.2).

In each group, individuals who participate in the government sector have compared to non-workers, on average, unobservable characteristics which lead to lower earnings in that sector. This may indicate that many low earnings workers in the government sector are forced to hold onto their jobs out of extreme need (i.e. they have very low reservation earnings). The pattern of sectoral self-selection found for urban men and women indicates that many of the urban workers who would be in the upper part of the private sector earnings distribution are remaining employed in the government sector, where they are also in the upper part of the earnings distribution.[24] While these workers are expected to receive higher earnings in the private sector, they likely remain in the government sector because of greater job security, the availability of non-pecuniary entitlements, and the possibility of financial gains from rent-seeking behavior.

Once controlling for both sources of selection, earnings are found to be higher in the private sector for all groups except rural women. Allowing the returns to human capital to vary across sectors also strongly affects the results. The null hypothesis that returns are equal across sectors

[24] In this situation, individual employed in the government sector earn more in the government sector than an average worker would, and they would also earn more than an average worker if they were employed in the private sector, but their advantage is greater in the government sector. Vice versa, those employed in the private sector earn less than an average worker would, but would also earn less than an average worker in the government sector and this would be a greater disadvantage than in the private sector.

is rejected at the 1% significance level for all four groups. In all cases, university educated Russians are found to have larger sectoral earnings differentials than Russians without general secondary education. This difference ranges from 3.8 (13) log percentage points for urban (rural) men to 22.1 (11.1) log percentage points for urban (rural) women.

B) The Selection-Corrected Sector-Specific Earnings Equations

Tables 4 & 5 present the complete results from the government and private sector earnings functions estimated above in specification 6. These earnings functions allow returns to human capital to fully vary across sector and jointly control for both types of self-selection. The dependent variable is log earnings at an individual's main job in the particular sector. In each table, the regression coefficients are presented along with their bootstrapped standard error.[25] All models include a two part error term which is composed of a fixed sector-year component to account for possible business cycle effects and a typical idiosyncratic component.

As seen in table 4, the human capital variables have little consistent effect on government sector earnings. A university degree is estimated to increase earnings by 20 – 30% (significant at the 5% level for all groups but rural men) and general secondary education to increase earnings by 3 – 8% for all individuals (insignificant for all groups). Specialized secondary education and professional courses are also found to have a positive effect on earnings, especially for women. Ordinary and secondary vocational degrees are found to have a strong negative effect on earnings. This may occur because either the skills associated with these degrees are becoming

[25] The significance level of each variable is calculated using the percentile bootstrap method and does not directly depend on its estimated standard error. The bootstrapped standard errors are robust to both strata level and household level clustering and correlation among the error terms. Appendix C.II provides additional information on the bootstrap method.

obsolete or the individual degree holders are segregated into lower earning occupations. Age has a weak positive effect on earnings for rural men with maximum earnings occurring at 34. Marriage also has a limited positive effect for men raising earnings by 5 – 8% on average. No marriage penalty is found for women (however, married women are found to be 8% less likely to participate in the labor force). The fixed sector-year component of the error term and the residential location variables are jointly significant for all four groups. Overall, the results provide some support for the conventional wisdom that earnings in the government sector are not dependent on individual productivity and are more likely related to the status of an individual's occupation and the overall political importance of their specific employer.

Table 5 presents the estimates of log earnings equations for the private sector. As in the government sector, human capital variables do not have a consistent effect on earnings, with only a university degree and weight significantly affecting earnings for three of the four groups. It is likely that human capital (especially education) acquired during the Communist period is not transferable to the new private sector and that unobservable factors, such as business skills or networking, better explain individual earnings. A university degree is found to increase earnings by 40 – 50% for all individuals (significant at the 5% level for all groups but rural women). Specialized secondary education is also found to have a strong positive effect (30 – 40%) on earnings for women. Age is a significant predictor of female earnings with the earnings of rural women reaching a maximum at 24 & 43 with a minimum at 33, and the earnings of urban women reaching a maximum at 39. Interestingly, all married individuals receive a large earnings premium ranging from 7% for urban women to 37% for urban men. The fixed sector-year component of the error term and the residential location variables are again jointly significant for all four groups.

V) Sectoral Earnings Differentials Across Various Groups

This section examines the selection-corrected sectoral earnings differentials in greater detail to get a better picture of which individuals benefit from private sector employment and which have higher earnings in the government sector. Except when noted, the sectoral earnings differential for a particular group is defined as the expected log earnings in the private sector minus those in the government sector for the average individual in that group, where expected earnings are based on the systematic parts of the earnings equations estimated in specification 6 and presented in tables 4 & 5, and are unconditional on the actual work status of individuals in that group.

A) Decomposition of the Unadjusted Sectoral Earnings Gap

Solving (C1) & (C2) for the average worker, the unadjusted sectoral earnings gap (the sample difference in mean sectoral earnings) can be written as

$$\overline{y}^{pr} - \overline{y}^{gv} = (\overline{X}^{*pr}\beta^{pr} + \overline{\lambda}^{*pr}\beta_{\lambda}^{pr}) - (\overline{X}^{*gv}\beta^{gv} + \overline{\lambda}^{*gv}\beta_{\lambda}^{gv}). \quad (6)$$

and decomposed into two components,

$$\overline{y}^{pr} - \overline{y}^{gv} = \overline{X}^{*pr}(\beta^{pr} - \beta^{gv}) + ((\overline{X}^{*pr} - \overline{X}^{*gv})\beta^{gv} + (\overline{\lambda}^{*pr}\beta_{\lambda}^{pr} - \overline{\lambda}^{*gv}\beta_{\lambda}^{gv})). \quad (7)$$

The first term on the right-hand side of (7) measures the unconditional effect of private sector employment on the earnings of the average private sector worker and is calculated as the difference between the average private sector worker's expected earnings in the private sector and those in the government sector. The second term measures the overall effect self-selection has on sectoral earnings and reflects both the difference in the observed skills of the average worker in each sector and the sample selection bias resulting from the correlation between unobserved characteristics and the average worker's likelihood of employment in each sector.[26]

[26] This procedure can be inverted to decompose the unadjusted earnings gap into the unconditional differential for

The first term can be further divided into the unconditional effect of private sector employment on the overall average worker and a second component with reflects both the sectoral difference in the returns to observed skills and the skill difference between the overall average worker and the average private sector worker.

$$\bar{X}^{*pr}(\beta^{pr} - \beta^{gv}) = (\bar{X}^{*}(\beta^{pr} - \beta^{gv}) + (\bar{X}^{*pr} - \bar{X}^{*})(\beta^{pr} - \beta^{gv})), \quad (8)$$

Table 6 presents the results of this decomposition for the four sample groups. The unadjusted earnings gap is positive for all groups with the average private sector worker earning 20 – 30 log percentage points more than the average government sector worker. For all groups besides rural men, the unconditional effect of private sector employment on the average private sector worker is quite different than the unadjusted earning gap with sample selection bias accounting for the majority of this difference. As already discussed, controlling for selection bias decreases the estimated effect of private sector employment for rural women because of strong negative self-selection of participants into the government sector. Meanwhile, the opposite effect occurs for urban women and men due to negative (positive) self-selection of workers into the private (government) sector.

In each group, the average private sector worker's skills are rewarded less in the private sector than those of the overall average worker. This is further evidence that workers who would gain the most from private sector employment are remaining employed in the government sector. However, contrary to this finding, male private sector workers do appear to be slightly more

---

the mean government sector worker and the corresponding selection effect component. In this case, the difference in returns to skill is calculated using the skills of the average government sector work and the difference in observed skill is calculated using the private sector returns to skill. As this procedure is not symmetric, the results will differ.

skilled than government sector workers.[27] Overall, both male and female private sector workers in urban areas are estimated to benefit substantially from private sector employment. The picture is quite different in rural areas where men are estimated to benefit somewhat from private sector employment and women actually receive lower earnings in the private sector. While not explored in detail in this paper, it is likely that both private sector infrastructure (i.e. property rights, contract enforcement, taxation policy, and financial institutions) and employment opportunities differ substantially between rural and urban area.

B) Sectoral Earnings Differentials by Age & Education

Table 7 presents sectoral earnings differentials calculated for the average worker in sixty-four groups defined by four age categories, four educational levels, gender, and urbanicity. The results differ substantially across groups and even within educational levels. For rural and urban men, a particular age group benefits the most from private sector employment across all educational levels. These groups are rural men aged 18-29 and urban men aged 30-39. For workers without university education, in rural areas, the youngest have the highest earnings differentials (although they are still negative for rural women), while in urban areas, those in their 30s have the highest differentials. This same pattern prevails for university educated male workers, but aging increases the earnings differential for university educated female workers, as the largest differential is found for rural women aged 40 – 49 and urban women aged 50 – 54 in this education group. More educated workers, in all cases besides the oldest urban men, have higher earnings differentials with the differences being particular large for rural women aged 40

[27] Some caution should be taken in interpreting this result since the differences in observed skills are calculated using the government sector returns to skill.

– 54, urban women and rural men aged 18 – 29 & 50 – 54, and urban men aged 18 – 29.

Figure 1 illustrates the age distribution of sectoral earnings differentials for rural and urban women and men at three educational levels. Here, the differentials are calculated for the mean individual in each of the twelve groups defined by education, gender, and urbanicity, and vary within these twelve groups only due to the changing returns to age. The results are similar to those presented in table 8. For all groups, differentials increase rapidly from age 20 and reach their initial peak around age 24 for rural women and men, and age 28 for urban women and men. For rural men, the differential decreases from its peak rapidly and then stabilize around 10 – 25 (30 – 40) log percentage points from age 35 – 59 for workers without (with) university education. The earnings differential also decreases from its initial peak for urban men and rural women. For urban men, this drop-off is more gradual than for rural men, accelerates after age 50, and is monotonic throughout the remaining age distribution for all education levels. For rural women, this decrease only lasts until age 33, where the differential begins to increase again. While not monotonic, it increases gradual throughout the remaining age distribution, and reaches its overall peak at age 54. Lastly, for urban women, earnings differentials, once peaking at age 28, remain stable throughout at around 100 – 110 (110 – 120) log percentage points for workers without (with) university education.

Overall, for men, it appears that youth benefit the most from private sector employment with earnings differentials deceasing by 80% (40%) as rural (urban) men age from their mid-20's to mid-30's (mid-50's). For women, the picture is more complex. Differentials also initially peak for young women, but older women, especially those over 50, actually benefit the most from private sector employment. One explanation for the different relative benefits of private sector employment found for older women and men is that under the Soviet system women were strongly segregated into white collar occupations. Even a decade after transition, 66% of

mathematicians, 63% of computer programmers, 64% of doctors, 94% of accountants, 78% of economists, and 98% of bookkeepers are female. It is likely that private sector employment is lucrative for many women in these occupations who are often in their 40's and 50's.

C) Sectoral Earnings Differentials by Unobserved Skills

Figure 2 presents sectoral earnings differentials for eighty groups defined by earnings deciles, the two education levels, gender, and urbanicity. Decile ranks of sample earnings are used to proxy for the average level of individual unobserved skill in each group. In each case, the earnings differential for the average worker increases as we move up the earnings distribution. This indicates that higher skill workers are better rewarded in the private sector relative to the government sector than those at lower skill levels, and is in accordance with the widely held belief that small private sector firms in Russia are profit maximizing and pay workers according to their productivity, while government sector firms often face soft budget constraints and must constrain wages in order to simultaneously maximize employment and minimize their wage bill. University educated workers typically have larger earnings differentials than the average worker. For urban women and men, this is true across the entire earnings distribution with earnings differentials 10 – 20 (5 – 15) log percentage points higher for university educated women (men). In contrast, for urban women and men, larger differentials are only found for university educated workers at the bottom (1^{st} – 3^{rd} decile for both groups) and top (10^{th} decile for women, 8^{th} – 10^{th} decile for men) of the earnings distribution.

VI) Conclusions

This paper makes two important contributions to the growing research on labor market outcomes in transitional economies. First, it shows that controlling for the endogeneity of participation and sectoral choice, in most cases, significantly affects estimates of government and private sector earnings equations for Russian workers. The results in tables 3 – 6 strongly suggest that both

types of self-selection bias ordinary least square (OLS) estimates of returns to human capital in Russia. As most estimates of changes in overall and within group inequality (such as gender inequality) for Russia have relied on OLS regression methods, the work in this paper implies that care must be taken in interpreting these past results.

Second, it analyzes sectoral earnings differentials by age, education, and unobserved skills in order to increase our knowledge of which Russians benefit from private sector employment and which appear to be losers in the transition to a market economy. In summary, earnings are estimated to be greater in the private sector for the average rural and urban male and urban female worker. Rural women and older rural and urban men are found to be the most vulnerable to decreases in government sector employment, while university educated workers, young rural and urban men, and urban women are among those most likely to gain from increases in private sector development.

While this basic study of sectoral earnings differentials suggests several reasons why the benefits of private sector employment may differ across these various groups, future research would ideally expand upon the analysis in this paper by accounting for: (1) The relationship between individual occupational choice decisions and the returns to private sector employment; and (2) Differential access to private sector employment in so far as market infrastructure and opportunities for private sector employment differ across regions and communities. Unfortunately, the data necessary for adding these refinements to the model are difficult to acquire and joint models of sectoral and occupational choice which control for self-selection are difficult to identify. For these reasons these refinements are left to future research.

References

Adamchik, Vera A. and Arjun S. Bedi. 2000. "Wage Differentials between the Public and the Private Sectors: Evidence from an Economy in Transition." *Labour Economics* 7: 203-24.

Åslund, Anders. 1997. "Observations on the Development of Small Private Enterprises in Russia." *Post-Soviet Geography and Economics* 38. no. 4: 191-205.

Brainerd, Elizabeth. 1999. "Winner and Losers in Russia's Economic Transition." *American Economic Review* 88 (December). no. 5: 1094-116.

Deaton, Angus. 1997. *The Analysis of Household Surveys*. Baltimore: The Johns Hopkins University Press.

Dustmann, Christian and Arthur van Soest. 1998. "Public and Private Sector Wages of Male Workers in Germany." *European Economic Review* 42: 1417-41.

The Economist (London). 1998. "A Backbreaking Job." 25 July: 50.

Flanagan, Robert J. 1994. "Labor Market Responses to a Change in Economic System." In *Proceedings of the World Bank Annual Conference on Development Economics 1994*: 405-25.

Hall P. 1992. *The Bootstrap and Edgeworth Expansion*. New York: Springer-Verlag.

Heckman, James. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47 (January). no. 1: 153-62.

Heeringa, Steven G. 1997. "Russia Longitudinal Monitoring Survey Sample Attrition, Replenishment, and Weighting: Rounds V-VII." *University of Michigan Institute for Social Research* (March).

Lee, Lung-Fei. 1978. "Unionism and Wage Rates: A Simultaneous Equations Model with Qualitative and Limited Dependent Variables." *International Economic Review* 19 (June). no. 2: 415-33.

Lin, Justin Y. and Guofu Tan. 1999. "Policy Burdens, Accountability, and the Soft Budget Constraint." *AER Papers and Proceedings* 89 (May). no. 2: 426-31.

OECD. 1998. "Public Sector Table 1." *National Accounts*. Paris.

Ogloblin, Constantin G. 1999. "The Gender Earnings Differential in the Russian Transition Economy." *Industrial and Labor Relations Review* 52 (July). no. 4: 602-27.

Stelcner, Morton, Jacques van der Gaag and Wim Vijverberg. 1989. "A Switching Regression Model of Public-Private Sector Wage Differentials in Peru: 1985 – 86." *Journal of Human Resources 24*. no. 3: 545-59.

Stillman, Steven. 2000. "Labor Market Uncertainty and Private Sector Labor Supply in Russia." *Rand Working Paper.*

Tunali, Insan. 1986. "A General Structure For Models of Double-Selection and An Application to a Joint Migration/Earnings Process With Remigration." in *Research in Labor Economics* 8B. Edited by Ronald G. Ehrenberg: 235-82.

Van der Gaag, Jacques and Wim Vijverberg. 1988. "A Switching Regression Model For Wage Determinants in the Public and Private Sector of a Developing Country." *The Review of Economics and Statistics* : 244-52.

Table 1: Summary Statistics by Work Status and Sector

Variables (N = 9239, N*T = 21120)	Not Employed	Government	Private
Percent of Observations	27%	53%	20%
Real Earnings		1,260	1,727
Year is 1994	26%	29%	28%
Year is 1995	24%	27%	27%
Year is 1996	24%	23%	22%
Year is 1998	26%	21%	23%
Male	43%	49%	57%
Age	34	38	36
Height (cm)	167	168	169
Weight (kg)	69	73	73
Married	65%	78%	74%
Some General Secondary Education (U.S. High School)	93%	96%	96%
Complete General Secondary Education (normal academic standard)	66%	70%	70%
Ordinary Vocational Diploma (lowest academic standard)	13%	15%	15%
Secondary Vocational Diploma (low academic standard)	22%	24%	26%
Specialized Secondary Diploma (highest academic standard)	23%	33%	28%
Professional Course Diploma	27%	32%	37%
Institute / University Diploma	12%	23%	20%
# Children	0.77	0.87	0.84
# Elderly	0.34	0.26	0.26
# Other Adults	1.35	1.24	1.24
# Other Workers	0.93	0.91	0.86
Owns Home	50%	44%	47%
Amount of Land Owned (Hectares)	0.20	0.11	0.48
Real Household Assets	91,466	104,874	113,667
Other Income	406	261	273
Region is Moscow / St. Petersburg	7%	9%	12%
Northern / North Western	7%	8%	7%
Central / Central Black-Earth	17%	19%	17%
Volga-Vaytski / Volga Basin	17%	19%	13%
North Caucasian	18%	11%	15%
Ural	14%	16%	15%
West Siberian	11%	10%	10%
East Siberian / Far Eastern	11%	10%	11%
Urban	65%	72%	73%
Lives In Capital (Moscow) / Oblast Center (regional capital)	37%	42%	45%
Raion Center (major town)	36%	36%	34%
Other	26%	22%	21%
Site Has Employment Service	75%	80%	81%
Had Enterprises Closed	54%	55%	58%
Has A Social Welfare Office	79%	82%	83%
Can Own Private Land	75%	72%	72%
Site Has No Banks	8%	6%	5%
Site Only Has RFSB (Russian Federation Savings Bank)	21%	18%	17%
Has Branches Besides RFSB	71%	76%	79%
Can Withdraw Money Fast (1,000 Rubles by the Next Day)	73%	77%	79%
Has A Close Bank (Nearest Non - RFSB Bank Is < 1 Mile Away)	46%	49%	51%

Note: All Values in Real 1998 Moscow / St. Petersburg Rubles (1 USD ≡ 25 Real Rubles)

Table 2: Identification of the Earnings and Reduced Form Selection Equations

Variable Groups Included	Earnings Equations	Participation Equation	Sectoral Choice Equation
Individual Characteristics	•	•	•
Household Demographics		•	•
Household Characteristics		•	•
Residential Location	•	•	•
Community Employment		•	
Community Characteristics			•

Note: Individual characteristics include gender, height, weight, age, marital status, educational status, and the year; Household demographics include the number of children, elderly, other adults, and other workers; Household characteristics include whether a home is owned, the amount of land owned, the value of real household assets, and the amount of other household income; Residential location variables include region, urbanicity, and city status; Community employment variables include whether there is an employment service, whether government enterprises have been closed, and whether there is social welfare office; and Community characteristics include whether individuals can own private land, the availability of formal banking, and other bank characteristics.

Table 3: Sectoral Earnings Differentials with Various Model Restrictions

Specification	1	2	3	4	5	6
Rural Women - N*T = 1981, N*T(Gov) = 1544, N*T(Priv) = 437						
Average Worker	0.264	0.253	-0.183	0.236	-0.169	-0.165
No General Secondary Ed.	0.264	0.297	-0.200	0.279	-0.186	-0.217
General Secondary Ed.	0.264	0.251	-0.194	0.227	-0.182	-0.167
University Diploma	0.264	0.187	-0.138	0.194	-0.116	-0.106
P - Value: $\beta^{gv} = \beta^{pr}$	0.006	NA	NA	NA	NA	NA
ρ_e: Corr (Participation, Sectoral Choice)	NA	NA	NA	NA	NA	0.737**
ρ_1^{gv}: Corr (Participation, Gov't Earnings)	NA	NA	-0.542**	NA	-0.540*	-0.470*
ρ_1^{pr}: Corr (Participation, Private Earnings)	NA	NA	0.120	NA	0.118	0.751
ρ_2^{gv}: Corr (Sectoral Choice, Gov't Earnings)	NA	NA	NA	0.128	0.033	-0.261
ρ_2^{pr}: Corr (Sectoral Choice, Private Earnings)	NA	NA	NA	-0.027	-0.018	0.048
Urban Women - N*T = 5516, N*T(Gov) = 4128, N*T(Priv) = 1388						
Average Worker	0.295	0.306	-0.087	1.536	1.156	0.973
No General Secondary Ed.	0.295	0.297	-0.144	1.519	1.081	0.912
General Secondary Ed.	0.295	0.278	-0.145	1.495	1.076	0.906
University Diploma	0.295	0.363	0.056	1.620	1.349	1.133
P - Value: $\beta^{gv} = \beta^{pr}$	0.000	NA	NA	NA	NA	NA
ρ_e: Corr (Participation, Sectoral Choice)	NA	NA	NA	NA	NA	0.741
ρ_1^{gv}: Corr (Participation, Gov't Earnings)	NA	NA	-0.420	NA	-0.394	-0.454
ρ_1^{pr}: Corr (Participation, Private Earnings)	NA	NA	0.320	NA	0.408	0.642
ρ_2^{gv}: Corr (Sectoral Choice, Gov't Earnings)	NA	NA	NA	-0.462*	-0.438*	-0.481
ρ_2^{pr}: Corr (Sectoral Choice, Private Earnings)	NA	NA	NA	-0.533	-0.551	-0.346
Rural Men - N*T = 2330, N*T(Gov) = 1592, N*T(Priv) = 738						
Average Worker	0.232	0.219	0.030	0.013	0.289	0.334
No General Secondary Ed.	0.232	0.179	-0.020	-0.021	0.252	0.297
General Secondary Ed.	0.232	0.220	0.031	0.027	0.304	0.345
University Diploma	0.232	0.363	0.205	0.087	0.362	0.427
P - Value: $\beta^{gv} = \beta^{pr}$	0.002	NA	NA	NA	NA	NA
ρ_e: Corr (Participation, Sectoral Choice)	NA	NA	NA	NA	NA	0.598
ρ_1^{gv}: Corr (Participation, Gov't Earnings)	NA	NA	-0.558**	NA	-0.399*	-0.287
ρ_1^{pr}: Corr (Participation, Private Earnings)	NA	NA	-0.232	NA	-0.351	-0.289
ρ_2^{gv}: Corr (Sectoral Choice, Gov't Earnings)	NA	NA	NA	0.601	0.410	0.285
ρ_2^{pr}: Corr (Sectoral Choice, Private Earnings)	NA	NA	NA	-0.172	-0.267	-0.341
Urban Men - N*T = 5565, N*T(Gov) = 3864, N*T(Priv) = 1701						
Average Worker	0.208	0.202	-0.039	1.166	0.808	0.972
No General Secondary Ed.	0.208	0.255	-0.034	1.263	0.848	1.007
General Secondary Ed.	0.208	0.183	-0.059	1.095	0.736	0.901
University Diploma	0.208	0.172	-0.010	1.165	0.879	1.045
P - Value: $\beta^{gv} = \beta^{pr}$	0.000	NA	NA	NA	NA	NA
ρ_e: Corr (Participation, Sectoral Choice)	NA	NA	NA	NA	NA	0.659**
ρ_1^{gv}: Corr (Participation, Gov't Earnings)	NA	NA	-0.806***	NA	-0.808***	-0.655**
ρ_1^{pr}: Corr (Participation, Private Earnings)	NA	NA	-0.423*	NA	-0.286	-0.476
ρ_2^{gv}: Corr (Sectoral Choice, Gov't Earnings)	NA	NA	NA	0.113	0.133	-0.118
ρ_2^{pr}: Corr (Sectoral Choice, Private Earnings)	NA	NA	NA	-0.581***	-0.558***	-0.606***

Note: The sectoral earning differential for group i is calculated as the expected log earnings in the private sector minus those in the government sector for the mean individual in group i. The underlying earnings equations are available by request from they author. The significance level of the error terms are calculated via the percentile bootstrap method, except of ρ_e which is calculated via a likelihood ratio test, and are robust to both strata level and household level clustering and correlation among the error terms.

* Significant at the 10 percent level. ** Significant at the 5 percent level. *** Significant at the 1 percent level.

Table 4: Selection-Corrected Log Earnings Equation for Government Sector Workers

	Female		Male	
	Rural	Urban	Rural	Urban
	Estimates of Coefficients and Bootstrapped Standard Errors			
Age	-6.010	2.266	-4.150**	-1.388
	(4.150)	(2.737)	(2.243)	(2.126)
Age^2 / 100	34.126	-14.607	22.901**	5.633
	(24.405)	(15.744)	(12.669)	(11.583)
Age^3 / 1,000	-9.407	4.514	-6.025**	-1.038
	(6.997)	(4.436)	(3.461)	(3.073)
Age^4 / 100,000	12.637	-6.697	7.590**	0.819
	(9.789)	(6.112)	(4.584)	(3.972)
Age^5 / 10,000,000	-6.645	3.829	-3.681**	-0.179
	(5.355)	(3.293)	(2.361)	(2.004)
Married	-0.026	0.048	-0.009	0.075*
	(0.095)	(0.065)	(0.147)	(0.114)
Some General Secondary Education	-0.293	-0.123	0.133	0.066
	(0.401)	(0.186)	(0.183)	(0.142)
Complete General Secondary Education	0.029	0.076	0.054	0.064
	(0.097)	(0.057)	(0.083)	(0.062)
Ordinary Vocational Diploma	-0.019	-0.085	-0.085	-0.135
	(0.106)	(0.074)	(0.087)	(0.073)
Secondary Vocational Diploma	-0.118	-0.073	-0.040	-0.112**
	(0.088)	(0.064)	(0.085)	(0.056)
Specialized Secondary Diploma	0.066	0.086**	0.051	0.021
	(0.108)	(0.065)	(0.133)	(0.067)
Professional Course Diploma	0.194*	0.139**	-0.191**	-0.047
	(0.111)	(0.055)	(0.099)	(0.057)
Institute / University Diploma	0.316**	0.256***	0.275	0.218***
	(0.135)	(0.083)	(0.152)	(0.069)
Height (cm)	0.006	-0.002	0.017**	0.002
	(0.007)	(0.004)	(0.007)	(0.005)
Weight (kg)	0.004	0.003*	-0.002	0.007***
	(0.003)	(0.002)	(0.004)	(0.003)
λ_1: Mills Ratio (Participation)	-0.633*	-0.616	-0.442	-1.084**
	(0.320)	(0.396)	(0.351)	(0.335)
λ_2: Mills Ratio (Sectoral Choice)	-0.352	-0.653	0.440	-0.195
	(0.504)	(0.372)	(0.481)	(0.369)
ρ_e: Corr (Participation, Sectoral Choice)	0.737**	0.741	0.598	0.659**
	(0.446)	(0.641)	(0.615)	(0.273)
ρ_1: Corr (Participation, Earnings)	-0.470*	-0.454	-0.287	-0.655**
	(0.220)	(0.285)	(0.228)	(0.186)
ρ_2: Corr (Sectoral Choice, Earnings)	-0.261	-0.481	0.285	-0.118
	(0.347)	(0.253)	(0.301)	(0.220)
σ_u: S.E. of the Regression	1.348	1.356	1.542	1.656
	(0.202)	(0.173)	(0.318)	(0.304)
Observations	1544	4128	1592	3864
R – Squared	0.0865	0.0522	0.0825	0.0727

Note: Jointly estimated reduced form selection equations are used to derive λ_1, λ_2, and ρ_e, and are available by request from the author. Each earnings equation includes a constant, residential location variables, and a two part error term which is composed of a fixed sector-year component to account for possible business cycle effects and a typical idiosyncratic component. The estimated standard errors are derived via bootstrapping, and are robust to both strata level and household level clustering and correlation among the error terms. The significance level of each variable is calculated using the percentile bootstrap method (except of ρ_e which is calculated via a likelihood ratio test) and does not directly depend on the estimated standard error .

* Significant at the 10 percent level. ** Significant at the 5 percent level. *** Significant at the 1 percent level.

Table 5: Selection-Corrected Log Earnings Equation for Private Sector Workers

	Female		Male	
	Rural	Urban	Rural	Urban
	Estimates of Coefficients and Bootstrapped Standard Errors			
Age	13.240*	7.608**	2.913	0.382
	(5.738)	(4.094)	(3.401)	(2.797)
Age^2 / 100	-76.986*	-41.838*	-14.251	-2.791
	(34.179)	(23.869)	(19.284)	(15.470)
Age^3 / 1,000	21.734	11.370*	3.325	0.905
	(9.888)	(6.806)	(5.283)	(4.161)
Age^4 / 100,000	-29.840	-15.256*	-3.761	-1.359
	(13.913)	(9.483)	(7.011)	(5.449)
Age^5 / 10,000,000	15.976	8.078*	1.668	0.762
	(7.632)	(5.168)	(3.618)	(2.784)
Married	0.194	0.067	0.258	0.374**
	(0.160)	(0.094)	(0.172)	(0.112)
Some General Secondary Education	0.402	0.596*	0.281	-0.488**
	(0.443)	(0.366)	(0.274)	(0.201)
Complete General Secondary Education	0.026	-0.003	0.041	-0.035
	(0.181)	(0.087)	(0.128)	(0.084)
Ordinary Vocational Diploma	0.185	-0.068	-0.083	-0.072
	(0.183)	(0.122)	(0.132)	(0.093)
Secondary Vocational Diploma	-0.043	-0.127*	0.074	0.010
	(0.147)	(0.091)	(0.129)	(0.078)
Specialized Secondary Diploma	0.390**	0.290**	-0.113	0.052
	(0.150)	(0.105)	(0.163)	(0.082)
Professional Course Diploma	0.077	-0.028	-0.004	0.043
	(0.143)	(0.089)	(0.125)	(0.073)
Institute / University Diploma	0.423	0.507***	0.440**	0.409***
	(0.236)	(0.130)	(0.187)	(0.096)
Height (cm)	0.018*	0.002	0.007	-0.014**
	(0.011)	(0.006)	(0.010)	(0.006)
Weight (kg)	-0.007	0.005**	0.008*	0.020***
	(0.005)	(0.003)	(0.005)	(0.003)
λ_1: Mills Ratio (Participation)	0.929	0.991	-0.377	-0.706
	(0.477)	(0.591)	(0.495)	(0.531)
λ_2: Mills Ratio (Sectoral Choice)	0.060	-0.535	-0.445	-0.899***
	(0.336)	(0.473)	(0.369)	(0.278)
ρ_e: Corr (Participation, Sectoral Choice)	0.737**	0.741	0.598	0.659**
	(0.446)	(0.641)	(0.615)	(0.273)
ρ_1: Corr (Participation, Earnings)	0.751	0.642	-0.289	-0.476
	(0.399)	(0.375)	(0.361)	(0.354)
ρ_2: Corr (Sectoral Choice, Earnings)	0.048	-0.346	-0.341	-0.606***
	(0.274)	(0.250)	(0.257)	(0.163)
σ_u: S.E. of the Regression	1.238	1.543	1.307	1.484
	(0.172)	(0.468)	(0.210)	(0.122)
Observations	437	1388	738	1701
R – Squared	0.1244	0.0887	0.1430	0.1125

Note: Jointly estimated reduced form selection equations are used to derive λ_1, λ_2, and ρ_e, and are available by request from the author. Each earnings equation includes a constant, residential location variables, and a two part error term which is composed of a fixed sector-year component to account for possible business cycle effects and a typical idiosyncratic component. The estimated standard errors are derived via bootstrapping, and are robust to both strata level and household level clustering and correlation among the error terms. The significance level of each variable is calculated using the percentile bootstrap method (except of ρ_e which is calculated via a likelihood ratio test) and does not directly depend on the estimated standard error .

* Significant at the 10 percent level. ** Significant at the 5 percent level. *** Significant at the 1 percent level.

Table 6: Decomposition of the Unadjusted Sectoral Earnings Gap

	Female		Male	
	Rural	Urban	Rural	Urban
Unadjusted Mean Sectoral Earnings Gap	.218	.245	.293	.213
I) Unconditional Differential for the Mean Private Sector Worker	-.207	.894	.323	.921
a) Unconditional Differential for the Overall Mean Worker	-.165	.973	.334	.972
b) Skill / Return Difference between the Overall Average Worker and the Average Private Sector Worker	-.042	-.079	-.011	-.051
II) Selection Effects	.425	-.649	-.029	-.707
a) Difference in Observed Skills	.003	-.018	.078	.031
b) Sample Selection Bias	.422	-.631	-.107	-.738

Note: The unadjusted mean earnings gap is the sample difference in the log earnings of the average private sector worker minus those of the average government sector worker. The unconditional differentials are calculated as the expected log earnings in the private sector minus those in the government sector. The skill / return difference is the effect of private sector employment on the overall average worker minus the effect on the average private sector worker. The difference in observed skills is calculated using the government sector returns to skill. Sample selection bias is then the residual component.

Table 7: Sectoral Earnings Differentials by Age and Education

	Average Worker	No General Secondary	General Secondary	University Diploma
Educational Group		Rural Women		
Age: 18 – 29	-.121	-.086	-.146	-.060
30 – 39	-.278	-.280	-.299	-.226
40 – 49	-.056	-.155	-.030	.013
50 – 54	-.245	-.527	-.053	-.016
		Urban Women		
Age: 18 – 29	.811	.745	.704	1.146
30 – 39	1.067	1.075	1.004	1.161
40 – 49	1.000	1.007	.937	1.090
50 – 54	1.003	.829	.979	1.172
		Rural Men		
Age: 18 – 29	.536	.485	.557	.762
30 – 39	.312	.299	.296	.407
40 – 49	.236	.219	.226	.357
50 – 59	.262	.214	.326	.386
		Urban Men		
Age: 18 – 29	.851	.850	.794	.992
30 – 39	1.082	1.155	1.002	1.164
40 – 49	1.034	1.099	.961	1.120
50 – 59	.873	1.046	.706	.805

Note: The sectoral earning differential for group i is calculated as the expected log earnings in the private sector minus those in the government sector for the mean individual in group i.

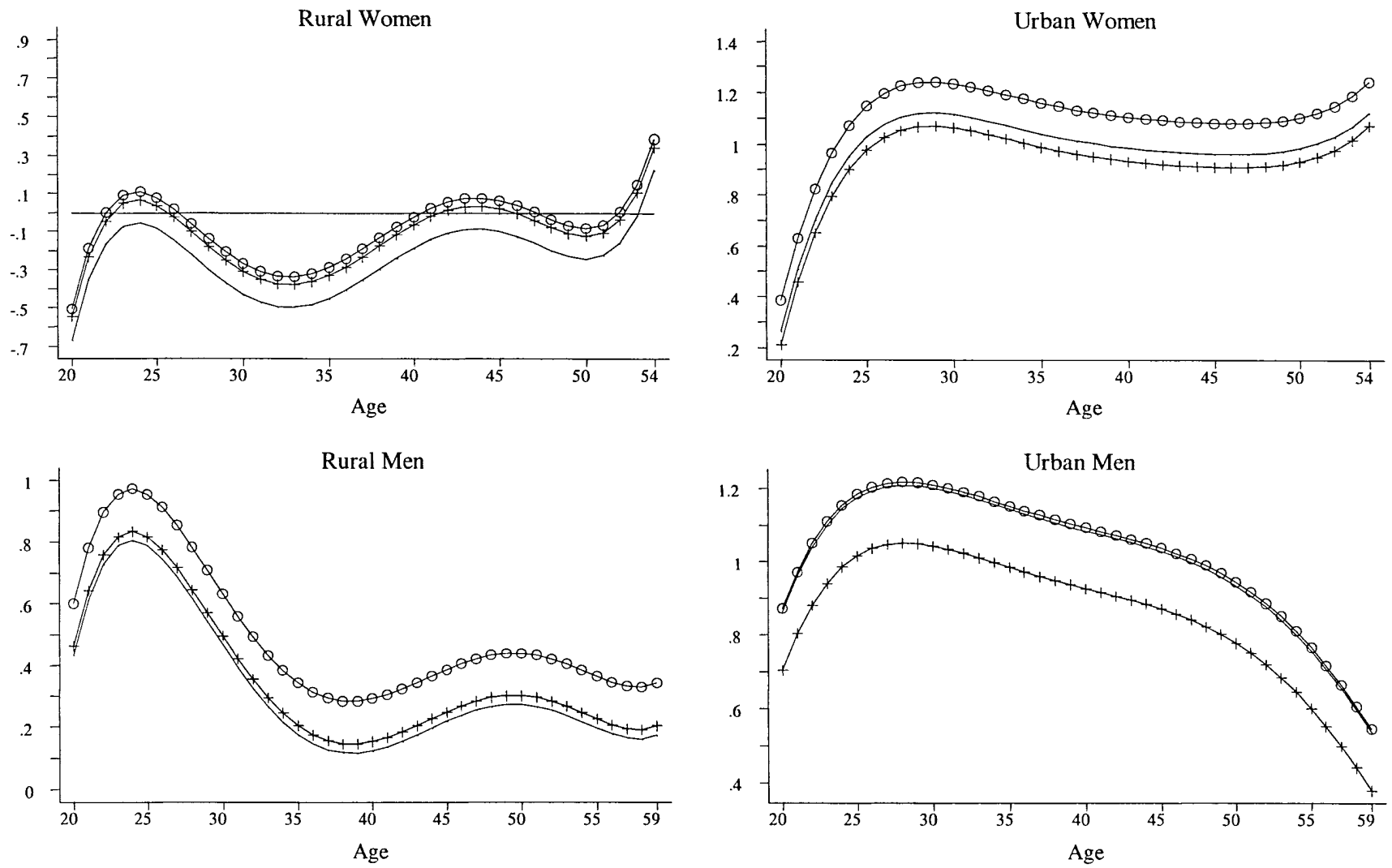

Figure 1: Sectoral Earnings Differentials by Age and Education

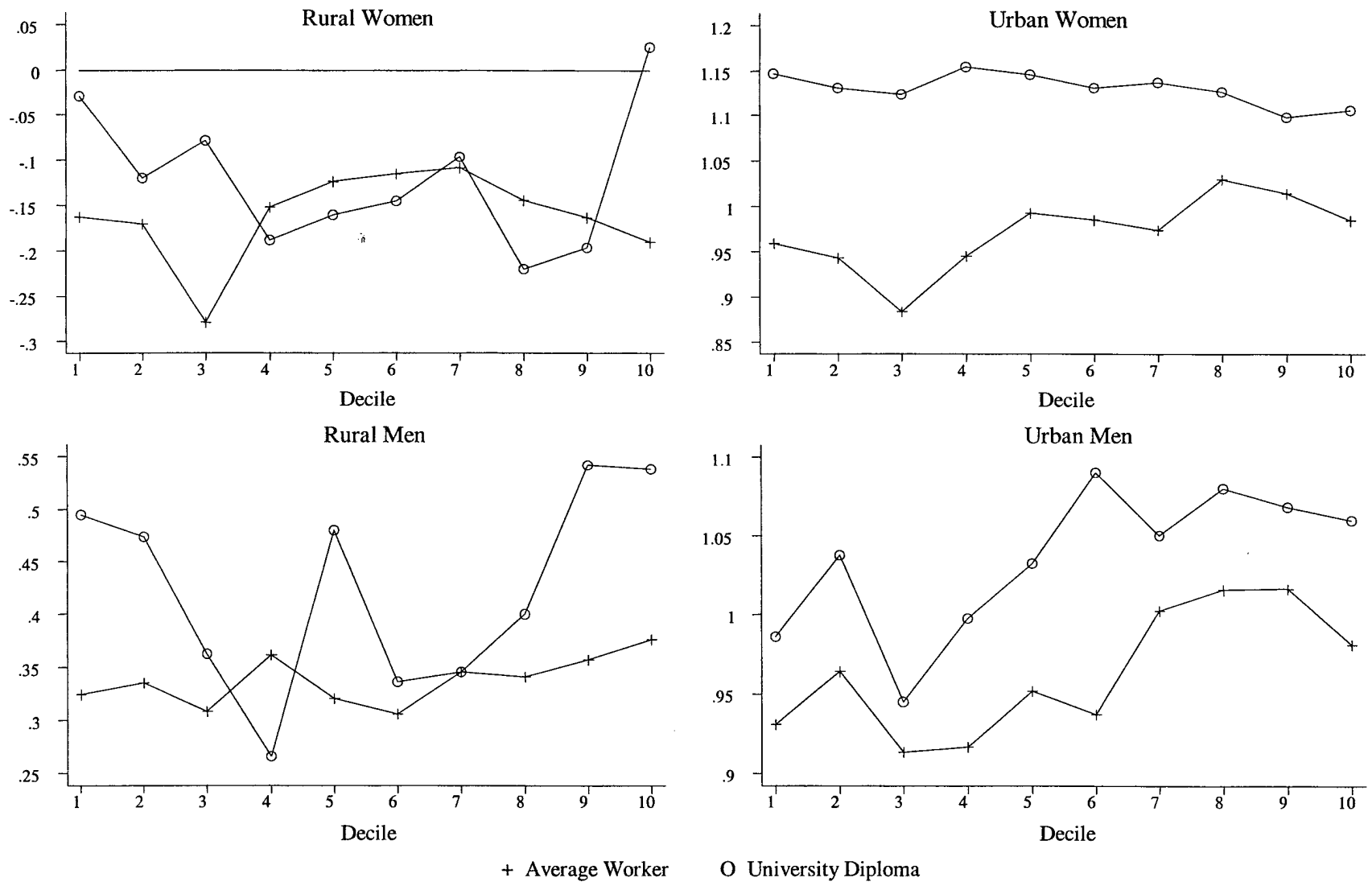

Figure 2: Sectoral Earnings Differentials by Earnings Decile

Appendix A: Summary Statistics

Table A1: Summary Statistics for the Entire Sample

Variables (N = 9239, N*T = 21120)	Mean	Min. / Max.	Standard Deviation
Individual Works	73%		0.44
Worker Is Employed In Private Sector	28%		0.45
Real Earnings For Workers in Government Sector	1,260	0 / 48,407	1,684
Real Earnings For Workers in Private Sector	1,727	0 / 56,672	2,816
Year is 1994	28%		0.45
Year is 1995	26%		0.44
Year is 1996	23%		0.42
Year is 1998	22%		0.42
Male	49%		0.50
Age	37	18 / 59	11
Height (cm)	168	113 / 201	9
Weight (kg)	72	29 / 190	14
Married	74%		0.44
Some General Secondary Education	96%		0.21
Complete General Secondary Education	69%		0.46
Ordinary Vocational Diploma	14%		0.35
Secondary Vocational Diploma	24%		0.43
Specialized Secondary Diploma	29%		0.46
Professional Course Diploma	32%		0.47
Institute / University Diploma	19%		0.40
# Children	0.84	0 / 8	0.95
# Elderly	0.28	0 / 3	0.55
# Other Adults	1.27	0 / 6	0.89
# Other Workers	0.91	0 / 6	0.77
Owns Home	46%		0.50
Amount of Land Owned (Hectares)	0.21	0 / 254	4.38
Real Household Assets	103,065	0 / 1,238,611	142,744
Other Income	302	-20,673 / 58,778	1,491
Region is Moscow / St. Petersburg	9%		0.29
Northern / North Western	7%		0.26
Central / Central Black-Earth	18%		0.38
Volga-Vaytski / Volga Basin	17%		0.38
North Caucasian	13%		0.34
Ural	15%		0.36
West Siberian	10%		0.30
East Siberian / Far Eastern	10%		0.30
Urban	70%		0.46
Lives In Capital / Oblast Center	41%		0.49
Raion Center	36%		0.48
Other	23%		0.42
Site Has Employment Service	79%		0.41
Had Enterprises Closed	55%		0.50
Has A Social Welfare Office	81%		0.39
Can Own Private Land	72%		0.45
Site Has No Banks	6%		0.24
Site Only Has RFSB	19%		0.39
Has Branches Besides RFSB	75%		0.43
Can Withdraw Money Fast	77%		0.42
Has A Close Bank	49%		0.50

Note: All Values in Real 1998 Moscow / St. Petersburg Rubles (1 USD ≡ 25 Real Rubles)

Appendix B: Sector of Employment

Individuals are asked detailed information about each of their jobs, which is used to classify the sector of employment. An individual is considered to work in the private sector at their main job if: (1) Their main job is other economic activity (such as sewing dresses, being a driver or delivery person, or doing household chores or repairs); (2) For their main job, they answered "no" to the question, "Do you work at an enterprise, organization, institution, collective farm, state farm, or firm?" and they are not in the army; (3) For their main job, they answered "yes" to the question, "What do you think, do you work at your own enterprise?" or (4) For their main job they answered "no" to the question "Is the government the owner or co-owner of your enterprise or organization?" they reported ≤ 200 workers at their place of employment, and they are not in the army. All individuals who answered these questions and who are not classified into the private sector are considered employed in the government sector at their main job. Not enough information is available to classify 7% of the main jobs. These are imputed by first using the job sector calculated in a previous or future year for individuals who report starting their current job before the year in question (56% of the imputations). Jobs which are still unclassified are then imputed using the predicted probabilities from a probit model which includes as independent variables; indicators of the individual's occupation, the year they started the job, the number of employees at the job, and the region and year of observation.[28]

[28] The percentage of individuals employed in the private sector in each region-year is used as the threshold value for the imputation.

Appendix C: Estimation Details

I) Estimating the Selection-Corrected Sector-Specific Earnings Equations

The entire reduced form model is estimated using the two-step procedure described in Tunali (1986).[29] The two reduced form selection equations (2) & (3) are estimated using a maximum likelihood bivariate probit model producing consistent estimates of γ, ψ, and ρ_e. Conditional on observability and dropping sector subscripts, expected log earnings in the private sector is

$$E\left[y_{it} \middle| W_{it} = 1, P_{it} = 1\right] = X^*_{it}\beta + b_{\lambda 1}\lambda_{1it} + b_{\lambda 2}\lambda_{2it}, \qquad \text{(C1)}$$

where $b_{\lambda k} = \rho_k \sigma_u$ for k = 1 & 2, $\lambda_{1it} = \dfrac{\phi(I_{1it})\Phi(I^*_{2it})}{\Phi^2(I_{1it}, I_{2it}, \rho_e)}$, $\lambda_{2it} = \dfrac{\phi(I_{2it})\Phi(I^*_{1it})}{\Phi^2(I_{1it}, I_{2it}, \rho_e)}$, $I_{1it} = Z^*_{1it}\gamma$,

$I_{2it} = Z^*_{2it}\psi$, $I^*_{1it} = \dfrac{I_{1it} - \rho_e I_{2it}}{(1-\rho_e^2)^{1/2}}$, $I^*_{2it} = \dfrac{I_{2it} - \rho_e I_{1it}}{(1-\rho_e^2)^{1/2}}$, and ϕ, Φ, and Φ^2 are, respectively, the standard normal density, the standard normal distribution, and the standard bivariate normal distribution function. Similarly, conditional on observability, expected log earnings in the government sector is

$$E\left[y_{it} \middle| W_{it} = 1, P_{it} = 0\right] = X^*_{it}\beta + b_{\lambda 1}\lambda_{1it} + b_{\lambda 2}\lambda_{2it}, \qquad \text{(C2)}$$

where $\lambda_{1it} = \dfrac{\phi(I_{1it})\Phi(-I^*_{2it})}{\Phi^2(I_{1it}, -I_{2it}, -\rho_e)}$, $\lambda_{2it} = \dfrac{-\phi(I_{2it})\Phi(I^*_{1it})}{\Phi^2(I_{1it}, -I_{2it}, -\rho_e)}$. The parameters of the sector-specific log earnings equations, β & $b_{\lambda k}$, are consistently estimated by computing $\hat{\lambda}_{kit}$ for each observation in the sample and estimating the sectoral earnings equations (1) by least squares

[29] This is an extended version of the two-step procedure used to estimate single selection switching models as original described in Heckman (1979) and Lee (1978).

regression of y_{it} on X_{it}^{*} and $\hat{\lambda}_{kit}$.[30]

It is also possible to obtain consistent estimates of the parameters σ_u and ρ_k for each sector. A consistent estimator of the standard error in each sector is found by: (1) Using the estimates of I_{1it}, I_{2it}, $b_{\lambda 1}$, $b_{\lambda 2}$, λ_{1it}, λ_{2it}, and ρ_e to calculate $\overline{\Omega} = \frac{1}{n}\sum_{it}\hat{\Omega}_{it}$, where $\hat{\Omega}_{it}$ is a complicated expression of all of these parameters, which is found in Tunali (1986; p. 276); (2) Calculating the least square residual variance, $\overline{\mathrm{T}} = \frac{1}{n}\sum_{it} u'u$, from each log earnings equation; and (3) Computing $\hat{\sigma}_u = \sqrt{\overline{\mathrm{T}} - \overline{\Omega}}$ for each sector. Consistent estimates for the ρ_k's in both sectors are now calculated as $\hat{\rho}_k = \hat{b}_{\lambda k} / \hat{\sigma}_u$.

II) Bootstrap Standard Errors & Confidence Intervals

As noted in Tunali (1986) & Heckman (1979), the estimated covariance matrices for the parameters in the selection-corrected sector-specific earnings equations are incorrect both because the error terms in earnings equations are heteroscedastic and because previously estimated values (the • 's) are used in estimating the earnings equations. While Tunali provides a procedure for generating appropriate asymptotic covariance matrices, this procedure relies on the delta method approximation and is extremely complicated.[31] Instead, this paper uses a

---

[30] Thus, the two λ's in each earnings equation are the double selection analogs of the inverse Mill's ratios included in single selection switching models.

[31] Furthermore, due to the clustered and panel nature of the dataset, observations are not independent across individuals, within households, or within stratas. To my knowledge it is not possible to adjust the formula for calculating asymptotic covariance matrices to account for this dependence between observations. However, as discussed in the next footnote, bootstrapping can be used to calculate correct standard errors for all parameters while

bootstrapping method to calculate correct standard errors for the selection-corrected earning equations. Bootstrapping, in general, can be used to recover the distribution of any defined statistic by exploiting the fact that the sample observations are a random sample of the overall population (Hall 1992). Thus, any random sample (with replacement) of the observed sample will also be a random sample of the overall population. The bootstrap method proceeds as follows: (1) A random sample of size N (all observations) is drawn with replacement from the observed sample; (2) The statistics of interest are calculated; (3) The process is repeated q times, with the value of the statistics noted each time; and (4) Given these bootstrapped estimates of the sampling distributions of the statistics, standard errors and confidence intervals are constructed. Accordingly, the standard errors for the coefficients in the sectoral earnings equations are calculated by first constructing 1000 bootstrap samples and then estimating the entire two-step model on each of these datasets, noting the resulting coefficients. The distribution of each coefficient is then used to calculate its correct standard error and confidence interval.[32]

---

accounting for this dependence.

[32] One assumption necessary with bootstrapping is the independence of observations in the dataset. In general, all datasets which use complex survey sampling methods violate this assumption (Deaton 1997; p.60). This paper makes two modifications to the simple bootstrap in order to apply it to the RLMS dataset. First, resampling is done independently within strata. Second, instead of resampling individuals, the procedure takes a weighted resample (weighted by the number of years the household is in the dataset) of the households in the dataset (splitoffs are included in their original households) and keeps all observations for a particular household when that household is selected. This procedure accounts for the observational dependence which is caused by both the clustered nature of the sampling procedure and the panel nature of the dataset.